Intuitive Biases in Choice versus Estimation: Implications for the Wisdom of Crowds
Author(s): Joseph P. Simmons, Leif D. Nelson, Jeff Galak, Shane Frederick
Source: *The Journal of Consumer Research*, (-Not available-), p. 000
Published by: The University of Chicago Press
Stable URL: http://www.jstor.org/stable/10.1086/658070
Accessed: 01/03/2011 14:11

Intuitive Biases in Choice versus Estimation: Implications for the Wisdom of Crowds

JOSEPH P. SIMMONS
LEIF D. NELSON
JEFF GALAK
SHANE FREDERICK

Although researchers have documented many instances of crowd wisdom, it is important to know whether some kinds of judgments may lead the crowd astray, whether crowds' judgments improve with feedback over time, and whether crowds' judgments can be improved by changing the way judgments are elicited. We investigated these questions in a sports gambling context (predictions against point spreads) believed to elicit crowd wisdom. In a season-long experiment, fans wagered over $20,000 on NFL football predictions. Contrary to the wisdom-of-crowds hypothesis, faulty intuitions led the crowd to predict "favorites" more than "underdogs" against point spreads that disadvantaged favorites, even when bettors knew that the spreads disadvantaged favorites. Moreover, the bias *increased* over time, a result consistent with attributions for success and failure that rewarded intuitive choosing. However, when the crowd predicted game outcomes by estimating point differentials rather than by predicting against point spreads, its predictions were unbiased and wiser.

Decades of research have uncovered the many ways in which consumers' judgments err (e.g., Alba and Hutchinson 2000; Bettman, Luce, and Payne 1998; Gilovich, Griffin, and Kahneman 2002; Kahneman and Tversky 2000; Simonson 1989; Thaler 1985), as well as the many ways in which consumers' judgments might be improved (e.g., Bertrand, Mullainathan, and Shafir 2006; Huber 1975; Thaler and Sunstein 2008). One of the more intriguing suggestions for improving judgments comes from a rapidly growing literature on the wisdom of crowds. The *wisdom-of-crowds hypothesis* predicts that the independent judgments of a crowd of individuals (as measured by any form of central tendency) will be relatively accurate, even when most of the individuals in the crowd are ignorant and error prone (Surowiecki 2004). Examples abound (Dunning 2007; Hastie and Kameda 2005; Lorge et al. 1958; Sunstein 2006; Surowiecki 2004; Yaniv 2004). Knight (1921) found that students' average estimate of the temperature of a classroom was only 0.4 degrees from accuracy, a result that was better than 80% of the individuals' judgments. Treynor (1987) asked 56 students to estimate the number of jelly beans in a jar. The average guess was 871, very close to the true number (850) and better than 98% of the students' individual guesses. Francis Galton (1907) reported the results of a regional fair competition that required people to estimate the weight of an ox. The average estimate was 1,197, just 1 pound away from the 1,198-pound ox's true weight!

The wisdom-of-crowds hypothesis has tremendous practical implications. First, it suggests that decisions made by majority rule (or by averaging opinions) will often outperform decisions made by single judges or experts (Hastie and Kameda 2005; Larrick and Soll 2006; Soll and Larrick 2009)

Joseph P. Simmons (joseph.simmons@yale.edu) is assistant professor of marketing, Yale School of Management, 135 Prospect Street, New Haven, CT 06520. Leif D. Nelson (leif_nelson@haas.berkeley.edu) is associate professor of marketing, Haas School of Business, University of California at Berkeley, Berkeley, CA 94720. Jeff Galak (jgalak@cmu.edu) is assistant professor of marketing, Tepper School of Business, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213. Shane Frederick (shane.frederick@yale.edu) is associate professor of marketing, Yale School of Management, 135 Prospect Street, New Haven, CT 06520. Address correspondence to Joseph P. Simmons. This research was partially funded by a Whitebox Behavioral Studies Fellowship Grant awarded to Joseph P. Simmons.

Baba Shiv served as editor and Gavan Fitzsimons served as associate editor for this article.

Electronically published December 2, 2010

 0093-5301/2011/3706-0006$10.00. DOI: 10.1086/658070

or decisions made by group discussion (Sunstein 2006). Second, it suggests that decisions made by majority rule (or by averaging opinions) will often be accurate in an absolute sense, an implication that partially accounts for the rapidly increasing use of information markets to forecast events and to inform policy decisions (Hahn and Tetlock 2006; Ho and Chen 2007). Indeed, as detailed below, crowd wisdom has been implicated as a cause of market efficiency (Surowiecki 2004; Treynor 1987).

Although researchers have documented many instances of crowd wisdom, it is important to go beyond these demonstrations to investigate whether some kinds of judgments may lead the crowd astray. Our investigation focuses on predictions made in a sports gambling context that features prices that are widely believed to reflect crowd wisdom. Contrary to popular belief, we find that knowledgeable crowds are typically *unwise* in this context because bettors rely on misleading intuitions about which team to bet on. We further show that the crowd's unwise reliance on intuition persists even when gamblers are warned that the intuitively appealing options are inferior and that this reliance *worsens* with feedback over time. Finally, we show that, although an explicit warning about the inferiority of intuitively appealing options fails to debias the crowd, we can debias crowd ignorance by eliciting predictions in a different, logically identical manner.

THE CONDITIONS OF CROWD WISDOM

The wisdom-of-crowds hypothesis derives from mathematical principles. If a crowd's judgment comprises signal-plus-noise, averaging judgments will cancel out the noise and extract the signal (Hogarth 1978; Makridakis and Winkler 1983).

Two conditions are necessary for the production of crowd wisdom. First, and most obviously, at least some members of the crowd must possess, and be motivated to express, relevant knowledge. For example, a crowd comprised entirely of people who know nothing at all about major league baseball would err considerably if its members were asked to predict the 2012 on-base percentage of Nick Markakis. Second, individual errors in judgment must not be systematic. For example, if all of the judges in a crowd make the *same* mistake, then averaging responses will obviously not negate the error. Because systematic errors compromise the production of crowd wisdom, it is important to identify conditions that decrease the likelihood of such errors. Scholars have emphasized two such conditions—independence and diversity. Independence is important because judges who talk to one another are likely to share the same knowledge and, hence, the same errors. Indeed, group discussion can reinforce or even exacerbate individuals' biases (Sunstein 2006). Similarly, diversity is important because even judges who do not interact may share the same knowledge (e.g., because they acquire information from the same sources) or desires and may therefore fall prey to the same errors. For example, salient but ill-founded rumors about a company's intention to acquire another company may influence the crowd's majority, and the crowd may consequently err in its assessment of the company's value (Shiller 2005).

In sum, wisdom-of-crowds proponents predict that crowds will be wise when the crowds' judges are (1) knowledgeable, (2) motivated to be accurate, (3) independent, and (4) diverse. The empirical question is whether this prediction is generally true.

Although most wisdom-of-crowds researchers have focused on documenting the surprising ability of crowds to make wise judgments, it is understood that crowds will perform poorly (relative to accuracy) when they are systematically biased. Thus, one threat to the generality of the wisdom-of-crowds hypothesis is the possibility that knowledgeable and motivated judges may systematically err even when the conditions of diversity and independence are met. Indeed, researchers in psychology, marketing, economics, and finance have spent decades documenting systematic biases in the ways in which individuals make judgments and decisions (e.g., Bettman et al.1998; Gilovich et al. 2002; Kahneman and Tversky 2000; Simonson 1989). For example, research shows that, on average, people are overly optimistic: they judge the outcomes of desirable events to be more likely than the outcomes of undesirable events (e.g., Forsythe, Rietz, and Ross 1999; Krizan and Windschitl 2007; Kunda 1990; Massey, Simmons, and Armor, forthcoming). Nevertheless, proponents of the wisdom-of-crowds hypothesis may find it easy to explain away this evidence. First, they may contend that many systematic biases arise only among populations (e.g., college students) that lack the requisite knowledge or only under conditions that provide no incentives for accurate responses. Second, systematic biases that persist even among the highly motivated and highly knowledgeable (e.g., optimism) may nevertheless produce errors that cancel out in a diverse sample (e.g., Camerer 1998). Thus, even if people tend to overestimate the likelihood of their preferred outcome, a crowd comprising people with different preferences may produce an average judgment that converges on the right answer. Finally, wisdom-of-crowds proponents may acknowledge that crowd wisdom will be compromised when judges are systematically biased but contend that such instances are rare in real market settings (List 2003).

Thus, a fair test of the wisdom-of-crowds hypothesis requires an investigation of a crowd of knowledgeable, independent, and diverse participants who have incentives to make accurate judgments in a realistic market setting. We report an experimental investigation that meets these requirements.

POINT SPREAD BETTING MARKETS

Point spread betting markets offer one of the most celebrated real-world examples of crowd wisdom (Surowiecki 2004) and one with enormous consequences. According to the American Gaming Association (2008), American consumers wagered $2.6 *billion* on sporting events in Nevada in 2007, and the AGA's Web site reports that this number represents less than 1% of all sports betting nationwide. By their estimate, Americans wager $380 billion on sporting events every year. That is more than the gross domestic product of Denmark.

To illustrate how these markets work, consider a National

Football League (NFL) game between the Baltimore Ravens and the Washington Redskins. At the time of this writing (and, frankly, throughout most of their history), the Ravens are vastly superior to the Redskins, and so, in the parlance of gambling, the Ravens would be the favorite and the Redskins would be the underdog. When gamblers bet on football or basketball games, they often do so against a point spread, a point amount that is subtracted from the favorite's score to better equate the two teams. A bet on the favorite wins only if the favorite wins by more than the point spread. A bet on the underdog wins if the favorite wins by fewer points than the spread or if the underdog wins the game. If the favorite's margin of victory is equal to the point spread, then the outcome of the bet is a tie and no money changes hands.

Many scholars and laypeople believe that point spreads are designed to generate equal betting on both teams (Avery and Chevalier 1999; Dana and Knetter 1994; Gandar et al. 1988; Gray and Gray 1997; Lee and Smith 2002; Oskarsson et al. 2009; Snowberg, Wolfers, and Zitzewitz 2005; Surowiecki 2004). For example, Dana and Knetter (1994) say that the spread "can be thought of as the best forecast of bettor behavior, rather than the best forecast of the game outcome" (1318), Avery and Chevalier (1999) report that "most accounts of the [point spread betting] market emphasize the propensity of casinos to set and alter the [spread] over time to balance betting" (502), and Snowberg et al. (2005) write that "half of the bets fall on either side" of the spread and that "the spread reveals the market's expectation of the median" outcome (367). According to this view, oddsmakers employed by casinos set an initial point spread, and bettors begin placing bets by deciding whether the favorite will win by more or less than the spread. Once the betting starts, oddsmakers adjust the point spread in an attempt to generate equal bets on both teams. For example, if most early bettors bet on the favorite, then the spread will be slightly increased in order to entice future gamblers to bet on the underdog. Because gamblers have to risk $11 in order to win $10, generating equal bets on each team guarantees a 5% profit for casinos, which are assumed to pursue this strategy in order to guarantee a profit and to avoid risking a loss on any of the games.

If point spreads generate equal bets on each team, then point spreads provide a reliable measure of the public's prediction of game outcomes. Surowiecki (2004, 13) writes: "A game's point spread ends up representing bettors' collective judgment of what the final outcome of that game will be." This is a belief that is implicitly or explicitly endorsed in many academic investigations of point spread markets (Avery and Chevalier 1999; Dana and Knetter 1994; Gandar et al. 1988; Gray and Gray 1997; Lee and Smith 2002; Snowberg et al. 2005). Moreover, if point spreads provide a measure of collective belief, then the accuracy of point spreads provides a measure of collective wisdom. And, in fact, because point spreads are extremely accurate (Radzevick and Moore 2008; Sauer 1998) and very difficult for gamblers to consistently defeat (Simmons and Nelson 2006), Surowiecki (2004, 13) has concluded that "the public . . . is pretty smart," and point spread accuracy is attributed to the emergence of crowd wisdom.

Unfortunately, this rosy conclusion is based on a false assumption. Point spreads do not, as is commonly believed, typically equate the bets on both teams (Jeffries and Oliver 2000; Levitt 2004; Roxborough and Rhoden 1998; Simmons and Nelson 2006), and therefore point spreads do not represent the crowd's prediction of game outcomes. In fact, point spreads rarely generate equal betting on each team, and so casinos often risk losing money on individual games. For example, in a sample of NFL football betting data that we scraped from Sportsbook.com's Web site in 2006 ($N =$ 192 games), we found that only 5.2% of the games featured a distribution of wagers that guaranteed a profit for the casino, meaning that the casino faced the possibility of a loss (and the possibility of a big win) in more than 94% of the games. Levitt (2004) similarly finds that "the bookmaker does not appear to be trying to set prices to equalize the amount of money bet on either side of a wager" (225). Casinos, it seems, adopt a long-term strategy that involves accepting losses on individual games while making a profit over a large sample of games.

As a consequence, point spread accuracy cannot be attributed to crowd wisdom but rather to the expertise of those who set and adjust the spreads. Thus, the efficiency of point spread betting markets is not indicative of crowd wisdom, and whether crowds are wise or unwise in these markets is a question that can only be answered by directly assessing the wisdom of gamblers.

INVESTIGATING CROWD WISDOM IN POINT SPREAD BETTING MARKETS

In point spread betting markets, wise crowds will predict without systematic bias and will choose wisely against inaccurate point spreads. However, past research suggests that wisdom may not prevail in this context (and in many contexts in which emotional, intuitive responses conflict with more rational, deliberative responses; Kahneman and Frederick 2002; Simmons and Nelson 2006). When predicting against point spreads, bettors' initial inclination—their intution—is to believe that the superior team (the favorite) will win against the spread. Moreover, bettors are usually quite confident in their intuition to choose the favorite—in Simmons and Nelson's (2006) parlance, most bettors have high *intuitive confidence*—and are therefore quite reluctant to abandon it. Thus, although point spreads quite accurately balance the favorite and the underdog, gamblers are substantially more likely to bet on favorites, and they seem to lend insufficient weight to point spreads when assessing which team is going to win against the spread (Levitt 2004; Simmons and Nelson 2006). Indeed, an analysis of predictions made by thousands of people competing in a fantasy football league found that the majority—the crowd—predicted favorites in over 90% of the games in their sample, even though favorites and underdogs were equally likely to win against the spread (Simmons and Nelson 2006).

Of course, when bets on favorites and underdogs are equally likely to win, betting on favorites more than underdogs does not constitute evidence that crowds are unwise, any more than would a systematic tendency to bet "tails" on a series of fair coin flips. Indeed, the wisdom-of-crowds hypothesis hinges on whether bettors predict accurately against inaccurate point spreads, such as those designed to exploit their judgmental tendencies. Indeed, although systematically betting on "tails" is at worst merely peculiar when the coin is fair, it is distinctly unwise if the coin systematically, and detectably, favors heads. Thus, it is important to know whether crowds bet on favorites more than underdogs even when point spreads are increased (and therefore biased against favorites). To date, only one study has investigated this question (Simmons and Nelson 2006, study 3b), and it did find, using a small sample of 10 games, that people predicted favorites more often than underdogs against increased spreads, a fact that decreased the accuracy of their predictions. However, this study did not provide an adequate test of the wisdom of crowds. Most notably, the study used a nondiverse, nonrepresentative sample of nonexperts (undergraduate football fans from Princeton University). Additionally, this study did not examine (1) whether this bias would emerge when participants *knew* that the point spreads disadvantaged favorites, (2) whether the bias would persist with feedback over time, and (3) whether eliciting predictions in a different manner would overcome this bias.

The experiment described below provides a rigorous and more comprehensive test of the wisdom-of-crowds hypothesis in this setting. We asked a knowledgeable and diverse sample of NFL football fans to predict NFL games against point spreads for the entirety of the 2007 17-week NFL season. The study's sample met all of the knowledge and diversity requirements suggested by wisdom-of-crowds proponents. And, critically, the point spreads were increased, allowing us to test whether crowds are appropriately sensitive to these increases, or whether they will wrongly choose favorites over underdogs the majority of the time.

Importantly, this experiment also allowed us to investigate three additional questions. First, although the strongest version of the wisdom-of-crowds hypothesis predicts that people will be sensitive to minor adjustments to the point spread even when they are not told of these adjustments (Surowiecki 2004), it is possible that a weaker version is more accurate—that people will abandon their intuitions and respond to point spread adjustments only if they are told that adjustments may have taken place. Indeed, if the crowd unwisely chooses underdogs because they simply do not realize that the favorites are disadvantaged, then warning them about this disadvantage should correct this mistake. In this experiment, we warned a randomly chosen subset of participants that many of the point spreads were increased, allowing us to test whether crowd wisdom increases when participants know that favorites are at a disadvantage on average.

Second, because the study was conducted over 17 weeks, this experiment gave us the opportunity to examine whether crowd wisdom improves with feedback over time. On the one hand, it seems sensible to expect crowds to get better over time, especially as bettors accumulate feedback suggesting that choosing favorites is unwise. On the other hand, some biases persist even in the presence of feedback (Massey et al., forthcoming) either because people fail to learn the correct rule (e.g., Gilovich 1983; people who are not told that the spreads disadvantage favorites may not learn that choosing favorites is unwise) or because they fail to apply a known rule to specific instances (e.g., Kahneman and Lovallo 1993; people may learn or know that favorites are generally an unwise choice, but the favorite may usually *seem* like the better bet whenever they consider specific games).

Third, this experiment investigated whether crowd wisdom depends on how predictions are elicited. In point spread betting markets, predictions are elicited by asking gamblers to choose which team to bet on against a given point spread, and when faced with this choice gamblers predict favorites more than underdogs (Levitt 2004; Simmons and Nelson 2006). But what if predictions were elicited not by asking people to choose against provided point spreads, but instead by directly asking people to estimate the point differential? On the one hand, the two questions are logically equivalent, and so one might expect them to elicit identical predictions. Indeed, when people predict that the favorite will win against a 10-point spread, they *should* also estimate the favorite to win by more than 10 points.

On the other hand, much research shows that logically identical methods of judgment elicitation can yield quite different judgments (Carmon and Simonson 1998; Fischer et al. 1999; Grether and Plott 1979; Shafir and LeBoeuf 2002; Slovic and Lichtenstein 1983; Tversky, Sattath, and Slovic 1988; Tversky and Thaler 1990). Such *preference reversals* emerge because different ways of asking the same question evoke different considerations and thought processes. In the point spread betting context, people who are making choices have to ask themselves: "Is the point spread big enough to convince me to abandon my intuition that the favorite is the right choice?" and because they are so confident in their intuitions, they underweight the point spread's magnitude and answer "no" to this question (Simmons and Nelson 2006). However, those generating estimates have only to ask themselves: "What will the point differential be?" That is a question that may focus them on the very dimension (the point spread) that typically receives insufficient weight when they are asked to decide which team to bet on. Thus, it is possible that although *choosing* against point spreads produces predictions that are biased in favor of intuition, estimating the point differential of each game produces predictions that are less biased—and therefore wiser.

In the experiment described below, we asked some of the participants to estimate the exact point differentials of each NFL football game, and they were rewarded based on how closely their prediction matched the eventual game outcome. This allowed us to test whether estimating exact point differentials yields less biased and more accurate predictions than choosing against point spreads and, thus, whether crowd wisdom can be improved by altering the way predictions are elicited.

RESEARCH QUESTIONS

To summarize, this research explored four research questions:

1. Are crowds wise when predicting against point spreads that disadvantage favorites?
2. Are crowds wise when they know they are predicting against point spreads that disadvantage favorites?
3. Does crowd wisdom improve over time?
4. Can crowd wisdom be improved by altering the way predictions are elicited?

THE EXPERIMENT

Participants

About 1 month before the start of the 2007 NFL football season, we recruited NFL fans to participate in a season-long NFL football study. We recruited participants by sending an e-mail to members of a Web site that we often use to conduct experiments, and we asked the members to forward the invitation on to NFL fans. People interested in participating followed a link to a Web page that asked them to provide their name, location (city and state), and favorite team. In addition, in an effort to identify knowledgeable NFL football fans, we asked them to rate how closely they followed the 2006 NFL football season (1 = not at all; 7 = extremely), and we asked them to recall, without looking up the answer, the two teams that played in the previous season's Super Bowl.

Over 1,000 people expressed interest in participating in the study, and more than 80% of them were not members of the Web site and thus were the by-product of word-of-mouth solicitations. We invited 240 people to register for the study a week before the start of the NFL season, 60 people for each of four experimental conditions. Of these, 178 (74.2%) did so, and only those who registered prior to the first week were invited to participate in subsequent weeks.

Although we did not advertise this fact, only those who reported following the previous NFL season "extremely closely" (i.e., a 7 on the scale) and who knew which teams played in the previous season's Super Bowl were deemed eligible for participation. Moreover, as indicated in an end-of-study survey, table 1 shows that our participants followed the 2007 NFL football season extremely closely and spent a great deal of time reading about and watching the NFL. Indeed, if one conservatively assumes that the average NFL game lasts 3 hours, table 1 shows that the median participant reported spending about 16 hours per week consuming NFL-related media. This is equivalent to about one-seventh of a typical person's nonsleeping hours.

Our sample was demographically and geographically diverse. Seventy percent of our participants were male, and their ages ranged from 18 to 60 years old, with an average age of 33. Our final sample of 178 participants lived in 40 different U.S. states (the most common state was California, the home of 9.6% of our sample). In addition, our sample had diverse rooting interests, as each of the 32 NFL teams was represented among the list of participants' favorite teams (the most common favorite team was the Pittsburgh Steelers, preferred by 11.2% of our sample).

TABLE 1

MEASURES OF NFL INVOLVEMENT

	Mean	Median
How closely you followed the season (9-point scale)	7.9	8.0
Number of games watched per week	3.6	3.0
Hours spent reading about the NFL per week	5.0	3.0
Hours spent watching NFL-related content per week	5.3	4.0
Number of games attended this season	1.0	0.0
Number of NFL jerseys you own	2.6	2.0

Experimental Conditions

At the start of the season participants were randomly assigned to one of four experimental conditions. They remained in their assigned condition for the duration of the experiment.

Participants assigned to the "choice" condition ($n = 43$) predicted NFL football games against point spreads that were increased relative to the official point spread. They were not told that the point spreads were increased. Participants in the "warned choice" condition ($n = 39$) faced an identical task, except that each week they were told that some of the point spreads were increased. Specifically, before making their predictions each week, they read: "Although official point spreads are designed to give each team an equal chance to win the bet, the point spreads inserted below are not necessarily the official point spreads. In fact, some of the point spreads have been increased, though none of them have been decreased. If you have read these instructions, please click the box below." Participants then clicked a box to indicate that they had read the warning.

Participants in the "estimate" condition ($n = 45$) did not make predictions against point spreads. Instead, they simply predicted which team would win the game and by how many points. Finally, participants in the "choice/estimate" condition ($n = 51$) predicted the winner of each game against a point spread before predicting the game's point differential.

Procedure

Participants logged on to a Web site each week to make their predictions. The Web site served as a home base for participants, who could use it not only to make predictions but also to check on the rules of participation, to review the terms of payment, to contact the experimenters, and to access their betting histories. A participant's betting history Web page displayed a list of every prediction he made, and it also kept an updated tally of the money he earned while participating in the study.

Every Thursday of the 17-week NFL football season, participants received an e-mail inviting them to make their predictions for the week. Although some NFL games were played on Thursdays, Saturdays, and Mondays, the vast majority of games were played on Sundays, and only Sundays featured at least one NFL game every week (in fact, it always featured at least 11 NFL games). Because of this, and to foster a weekly routine, we asked participants to predict "only" games played on Sundays. In total, they were asked to predict the outcomes of 226 games. They could submit their predictions until 1 hour before the first game of the week was scheduled to begin (their usual deadline was 12 p.m. Eastern Time on Sunday), and once their predictions were submitted they could not alter them. Participants who did not submit their predictions by Saturday afternoon were sent a reminder e-mail, and those who missed the deadline did not participate in that particular week of the study (but they were invited to participate in all subsequent weeks).

Each week, participants in the choice and warned choice conditions were presented with the list of games that would be played on Sunday. Each game listed the visiting team followed by the home team, and the point spread was provided in parentheses next to the favorite, as is customary. For example, a game played between the visiting Miami Dolphins and the home Washington Redskins appeared as, "Miami Dolphins at Washington Redskins (−4.0)." The "(−4.0)" was the point spread, meaning that a bet on the Redskins would win if the Redskins won by more than 4 points, and a bet on the Dolphins would win if the Redskins won by less than 4 points or if the Dolphins won the game. Before making their predictions in the first week of the season, participants in the choice and warned choice conditions underwent a tutorial to ensure that they understood the rules of predicting against point spreads, and they could access this tutorial via the study's Web site at any time. No participants ever questioned the rules or challenged their earnings, facts that strongly suggest that all participants understood the rules of predicting against point spreads.

The wisdom-of-crowds hypothesis predicts that crowds will take advantage of attempts to exploit them, thereby emerging "wise." To test this, the point spreads were adjusted in an attempt to exploit previously identified systematic tendencies. Most notably, because prior research found that people are more likely to bet on favorites than on underdogs (at least against accurate point spreads), we increased the point spreads for every game, thereby making underdogs more likely to win against the spread. In addition, prior research has found that people are more likely to bet on visiting favorites than on home favorites, presumably because people underestimate the NFL's significant home field advantage (Levitt 2004; Simmons and Nelson 2006). In an attempt to exploit this potential source of error, we increased the spreads a greater amount (usually three points) when the favorite was the visiting team than when it was the home team (usually one point). All spread adjustments were made on Wednesday evening of each week, and so those adjustments were based on official point spreads retrieved at that time from vegasinsider.com. Although slight changes to the official spreads between Thursday and Sunday were common—due to news of injuries, weather, and so forth (Roxborough and Rhoden 1998)—the spreads we provided were never altered once participants were invited to make their predictions on Thursday. This ensured that all participants made predictions against the same point spreads regardless of when their predictions were submitted.

Participants in the choice and warned choice conditions were asked to assign one of five possible wager amounts to each prediction: $0.50, $1.00, $1.50, $2.00, or $2.50. Requiring participants to wager at least $0.50 on each game helped ensure that they were motivated to accurately predict every game. Because we had a limited, though reasonably sized, budget for this study, the sum of participants' weekly wager amounts could not exceed an average of $1.50 per game. For example, participants could not wager more than $21.00 in a week featuring 14 games. Participants were instructed that, within each week, a winning bet would earn them the amount they wagered, a losing bet would lose them the amount they wagered, and a tie would earn them $0. Importantly, to help prevent systematic attrition, all weeks were independent and participants could never lose money by participating in this study. For example, if a given week resulted in $5.00 worth of winnings, then the participant won $5.00 and that was his to keep no matter how he performed during subsequent (or previous) weeks. If, however, a given week resulted in a total loss, then this was not treated as a loss but rather as a gain of $0. These rules were implemented to encourage participants to participate each week of the season no matter how much they had won or lost previously and no matter how much they expected to win or lose in the future. At the same time, these rules did not disturb participants' incentives to provide accurate predictions each week. Participants in all conditions received a gift certificate for the amount of their winnings at the end of the season.

Each week participants in the estimate condition were presented with the list of games to be played on Sunday, but the games were presented without point spreads. Thus, the Dolphins/Redskins game alluded to earlier was presented simply as "Miami Dolphins at Washington Redskins." For each game, participants first predicted which team would win the game and then they predicted how many points the chosen team would win by. Participants were paid based on how closely their prediction matched the actual game outcome. They were paid $2.50 for a perfect prediction, $2.00 for a prediction that deviated by one point, $1.50 for a prediction that deviated by two points, $1.00 for a prediction that deviated by three points, and $0.50 for a prediction that deviated by four points. Participants earned nothing for predictions that deviated by more than four points.

Finally, participants in the choice/estimate condition first made predictions against the same point spreads used in the choice and warned choice conditions. Then, as in the estimate condition, they predicted which team would win the game and by how many points. Although participants in this condition first made predictions against point spreads, they did not set wager amounts, and they were not compensated based on the accuracy of this prediction. Rather, their compensation was based solely on their point differential prediction, exactly as participants in the estimate condition were compensated. We included this condition to help us determine whether any differences that arose between the choice and estimate conditions were attributable to (1) merely considering the point spreads and/or (2) being asked to predict the exact point differential of the game.

Follow-Up Survey

Approximately 1 week after the season ended, participants were asked to complete an online follow-up survey, and 167 of the 178 original respondents did so. We constructed two versions of the survey—one for participants in the choice and warned choice conditions and one for participants in the estimate and choice/estimate conditions. For all participants, the survey asked them questions (shown in table 1) designed to assess their level of involvement in NFL football. All participants were also asked to rate their liking of each of the 32 NFL teams on a scale ranging from -3 (strongly dislike) to $+3$ (strongly like). Finally, only choice and warned choice participants (1) indicated whether they believed the point spreads were generally unbiased, too high, or too low and (2) completed a survey designed to assess whether they attributed winning or losing predictions to luck or to skill (described in more detail below).

RESULTS

Attrition

Attrition was minimal in this study. Of the 178 original participants, only 9 (5.1%) quit before week 10, and only 12 (6.7%) quit before week 14. The average participant made 202 predictions (89.4%), participated in 15.2 weeks, and quit 16.1 weeks into the 17-week season. Importantly, these measures of attrition did not differ by condition (p's > .65). Within the two estimate conditions, attrition did not correlate with earnings (p's > .18). Within the two choice conditions, attrition did correlate with earnings: participants who remained in the study longer tended to earn more money per prediction than participants who quit the study earlier (r's > .40, p's < .001). If one assumes that participants who remained in the study were better predictors of NFL football games than participants who quit, this pattern of attrition would increase the probability that the choice condition crowds would perform better over time. As reported below, this did not happen.

In sum, attrition in this study was very low, and it did not differ by condition. Because of this and because the results reported below are not affected by removing participants who quit the study early, we included all participants in the analyses reported below.

Are Crowds Wise When Choosing against Point Spreads?

According to the wisdom-of-crowds hypothesis, a majority of knowledgeable, motivated, independent, and diverse individuals will choose wisely when predicting against inaccurate point spreads even when they are not told that the point spreads are inaccurate. To test this hypothesis, we determined, for each game, the choice condition crowd's prediction against the spread. We did this in two different ways. By the "wager" method, we determined whether the percentage of money wagered on the favorite was greater than, less than, or equal to 50%, indicating a choice of "favorite," "underdog," or "no preference," respectively. This method gave greater weight to participants who bet more money on the game. By the "counting" method, we simply determined whether the percentage of people choosing the favorite was greater than, less than, or equal to 50%. This method gave each participant equal weight regardless of how much they wagered on the game. For simplicity, we will report only the results of the wager method except when the results of the two methods differ. However, results of the counting method are included in tables 2–6.

Because all of the point spreads in this study were increased, favorites lost more games than they won against the spread (98 wins, 124 losses, and 4 ties), and predicting favorites was therefore an unwise strategy. The wisdom-of-crowds hypothesis predicts that the choice condition crowd will (1) tend to choose underdogs more than favorites against increased point spreads, (2) win more games than it loses, and (3) outperform most of its individual members. None of these predictions was confirmed (see tables 2–4). In con-

TABLE 2

THE CROWD'S PREDICTIONS AGAINST THE SPREAD (N = 226 GAMES)

	Favorite	Underdog	No preference
Choice:			
Counting	198	23	5
	$(87.6)_{ab}$	(10.2)	(2.2)
Wager	202	24	0
	$(89.4)_{a}$	(10.6)	(.0)
Warned choice:			
Counting	188	34	4
	$(83.2)_{ab}$	(15.0)	(1.8)
Wager	187	38	1
	$(82.7)_{b}$	(16.8)	(.4)
Estimate:			
Median	59	159	8
	$(26.1)_{d}$	(70.4)	(3.5)
Mean	39	187	0
	$(17.3)_{e}$	(82.7)	(.0)
Choice/estimate:			
Median	95	123	8
	$(42.0)_{c}$	(54.4)	(3.5)
Mean	53	170	3
	$(23.5)_{de}$	(75.2)	(1.3)

NOTE.—Within the "Favorite" column, percentages (figures in parentheses) with different subscripts differ significantly ($p < .05$).

TABLE 3

THE CROWD'S PERFORMANCE AGAINST THE SPREAD (EXCLUDING TIES)

	Wins	Losses
Choice:		
Counting	93	124
	$(42.9)_{bc}$	(57.1)
Wager	96	126
	$(43.2)_{bc}$	(56.8)
Warned choice:		
Counting	88	130
	$(40.4)_{c}$	(59.6)
Wager	93	128
	$(42.1)_{bc}$	(57.9)
Estimate:		
Median	109	105
	$(50.9)_{ab}$	(49.1)
Mean	123	99
	$(55.4)_{a}$	(44.6)
Choice/estimate:		
Median	103	111
	$(48.1)_{ab}$	(51.9)
Mean	111	108
	$(50.7)_{ab}$	(49.3)

NOTE.—Within the "Wins" column, percentages (figures in parentheses) with different subscripts differ significantly ($p < .05$).

trast to the wisdom-of-crowds hypothesis, the choice condition crowd unwisely bet on the favorite in 89.4% of the games in the sample ($\chi^2(1, N = 226) = 140.19, p < .001$), thus exhibiting the same strong tendency to choose favorites found in research using unbiased spreads (Simmons and Nelson 2006). As a result, the choice condition crowd lost significantly more games (56.8%) than it won (43.2%; $\chi^2(1, N = 222) = 4.05$, $p = .044$), and the crowd performed worse than 93% of its individual members. Clearly, crowd wisdom was absent from this condition.

Although the strong version of the wisdom-of-crowds hypothesis predicts that crowds will predict wisely against inaccurate point spreads, a weaker version predicts that crowds will predict wisely only when they are told that the spreads are inaccurate. Consistent with this, the warned choice condition crowd, which was told that some of the spreads were increased, predicted slightly fewer favorites than the choice condition crowd: using the wager method ($\chi^2(1, N = 226) = 4.15$, $p = .041$; using the counting method, $\chi^2(1, N = 226) = 1.77$, $p = .182$). However, as shown in tables 2–4, the warned choice condition crowd also predicted favorites for the vast majority of the games (82.7%; $\chi^2(1, N = 226) = 96.92$, $p < .001$), also lost (57.9%) more games than it won (42.1%; $\chi^2(1, N = 221) = 5.54$, $p = .018$), and also performed worse than almost all (97.4%) of its individual members. Thus, the crowd was unwise even when its members were told that the spreads were increased.

The small effect of warning on predictions may have been due to the failure of participants in the warned choice condition to attend to or believe the warning. However, there are reasons to doubt this. First, we required all participants in the warned choice condition to check a box to indicate that they had read the warning, thus making it very difficult for them to ignore it completely. Second, favorites lost more often than they won in this study; thus, participants were exposed to feedback consistent with the warning, which should have increased their tendency to believe it. Third, in the end-of-season survey, most of the participants in the warned choice condition reported that the spreads were too high (65.7%) versus too low (5.7%) or unbiased (28.6%). Moreover, the tendency to believe that the spreads were too high was greater in the warned choice condition than in the choice condition (42.5%; $\chi^2(1, N = 75) = 4.04$, $p = .044$). This suggests that most participants attended to the warning and believed it. Finally, it is worth noting that the warned choice condition results presented in tables 2–4 are unchanged if we only include participants who reported that the spreads were "too high" in the end-of-season survey. Thus, even a crowd comprising solely those who acknowledged that the spreads were biased against favorites predicted favorites (78.7%) more than underdogs ($p < .001$), won (43.0%) fewer games than it lost ($p = .037$) and was outperformed by almost all (96.2%) of its individual members.

We have argued that the choice and warned choice crowds predicted more favorites than underdogs because they unwisely believed that favorites were more likely to beat the spread. Alternatively, it could be argued, the crowd may have wisely believed that favorites were less likely to win but chose to forgo financial gain in the service of a *preference* for betting on favorites. If this were true, then we would expect participants to have wagered less on favorite predictions than on underdog predictions in order to minimize their financial stake in an enjoyable but unwise selection. Alternatively, if participants were simply unwise in their assessment of the game outcome, then they should either have wagered an equal amount on predicted favorites and predicted underdogs, or, if they were (unwisely) more

TABLE 4

THE CROWD'S PERFORMANCE RELATIVE TO THE PERFORMANCE OF ITS INDIVIDUAL MEMBERS

	% of individuals predicting a higher percentage of favorites against the spread than the crowd	% of individuals the crowd outperformed
Choice:		
Counting	4.7	7.0
Wager	2.3	7.0
Warned choice:		
Counting	7.7	.0
Wager	10.2	2.6
Estimate:		
Median	80.0	57.8
Mean	91.1	95.6
Choice/estimate:		
Median	68.6	35.2
Mean	90.2	54.9

TABLE 5

PERCENTAGE OF GAMES THE CROWD PREDICTED FAVORITES AGAINST THE SPREAD

	Weeks 1–4	Weeks 5–8	Weeks 9–12	Weeks 13–17
Choice:				
Counting	82.1	87.2	90.2	97.0
Wager	80.4	85.4	94.3	95.7
Warned choice:				
Counting	81.8	77.1	88.0	89.9
Wager	75.0	79.2	86.8	89.7

confident in favorite than underdog predictions, they should have wagered more on predicted favorites than on predicted underdogs. In fact, participants wagered *more* money per predicted favorite than per predicted underdog (Simmons and Nelson 2006). Across games, choice condition crowd members wagered an average of \$1.45 on each favorite prediction and \$1.28 on each underdog prediction ($t(225) = 9.94$, $p < .001$). Warned choice condition crowd members wagered an average of \$1.48 on each favorite prediction and \$1.37 on each underdog prediction ($t(224) = 6.37$, $p < .001$). (We eliminated one game in which all warned choice members chose to bet on the favorite. There were no games for which everyone bet on the underdog.) These results favor the conclusion that crowds unwisely believed that favorites would win against the spread, and they challenge the conclusion that crowds defied their wisdom by knowingly betting on inferior but preferred options.

In sum, the results reported in this section fail to support the wisdom-of-crowds hypothesis. When they are asked to choose which team to bet on, crowds are not sensitive to point spread adjustments even when they are told that adjustments have been made. Instead, predictions against point spreads seem guided less by accuracy than by a reliance on intuition that causes people to bet on favorites more than underdogs, a systematic tendency that, in this circumstance, costs money.

Does Crowd Wisdom Increase over Time?

Although crowd wisdom was absent from each of the choice conditions, perhaps it simply takes some time for people to accrue enough experience to learn that favorites are losing more often than they are winning. Thus, it is reasonable to expect crowd wisdom to increase over time as evidence of the inferiority of favorites increases. Once again, however, the data fail to support this hypothesis. As shown in table 5, the tendency for the crowd to unwisely predict favorites actually increased over the course of the season. The correlation between time (the week predictions were made) and the tendency to predict favorites was positive and significant in the choice condition ($r(224) = .21$, $p = .002$) and the warned choice condition ($r(223) = .14$, $p = .034$). (The correlation in the warned choice condition was marginally significant [$p = .14$] when we analyzed predictions using the counting method.) The relationship between time and accuracy was negligible ($r(220) = .03$, $p = .696$ in the choice condition; $r(219) = .03$, $p = .679$ in the warned choice condition). Moreover, as shown in tables 5 and 6, the crowd was biased toward favorites throughout the season, and at no 4-week stretch did the crowd perform better than 50%. Thus, the crowd did not improve as it accrued knowledge and experience.

It is interesting to consider why the crowd increased its predictions of favorites over time despite the fact that favorites performed poorly against the spread. If the crowd was wise, this pattern should emerge only if the performance of favorites improved over time. However, the performance of favorites did not differ over time ($r(220) = .01$, $p = .905$), making wisdom an unlikely explanation. We will entertain two possibilities.

First, research shows that the more confidently people believe that the favorite will simply win the game, the more likely they are to predict favorites to win against the spread (following Simmons and Nelson [2006], we will refer to this belief as *intuitive confidence* because it represents confidence in the intuition that the favorite will win). Thus, one plausible explanation is that people became increasingly certain that favorites would win the games as the season progressed and that this increase in intuitive confidence led to an increase in betting on favorites. To investigate this possibility, we used data from Yahoo.com's fantasy football Web site. Each week of the football season thousands of people log onto Yahoo.com to compete in accurately predicting the winners of NFL football games, and each week Yahoo.com reports the percentage of people predicting each team to win. For each of the games in our sample, we measured intuitive confidence by recording the percentage of people in the Yahoo.com sample who predicted that the favorite would win the game. This reasonably assumes that games with a greater percentage of people believing that the favorite will win are associated with greater intuitive confidence (cf. Koriat 2008).

Consistent with Simmons and Nelson (2006), intuitive confidence strongly predicted when the crowd chose favorites ($r(224) = .42$, $p < .001$ in the choice condition; $r(223) = .42$, $p < .001$ in the warned choice condition [see table 7]). In addition, there was a trend for intuitive confidence to increase over time ($r(224) = .11$, $p = .102$). This indicates that as the season progressed, people were marginally more confident that the favorite would win the games. How-

TABLE 6

THE CROWD'S WINNING PERCENTAGE AGAINST THE SPREAD

	Weeks 1–4	Weeks 5–8	Weeks 9–12	Weeks 13–17
Choice:				
Counting	39.3	42.2	42.0	47.0
Wager	41.1	41.3	40.4	48.5
Warned choice:				
Counting	40.0	32.6	36.7	48.5
Wager	42.9	39.1	36.5	47.8

TABLE 7

PERCENTAGE OF GAMES IN WHICH THE CROWD PREDICTED FAVORITES AGAINST THE SPREAD

	Preferred favorite	Indifferent	Preferred underdog
Choice:			
Counting	92.2	86.0	68.4
Wager	89.2	80.1	66.7
Warned choice:			
Counting	86.0	80.3	70.0
Wager	80.1	73.2	66.2
Estimate:			
Median	37.5	27.8	20.6
Mean	36.0	20.4	18.7
Choice/estimate:			
Median	54.5	47.5	28.4
Mean	42.8	33.3	16.0

ever, although the relation between time and the warned choice condition's predictions of favorites decreased somewhat after controlling for intuitive confidence ($r(222) = .11$, $p = .110$), the choice condition's predictions of favorites increased over time even after controlling for intuitive confidence ($r(222) = .18$, $p = .007$). This increase in intuitive confidence may have contributed to the increase in betting on favorites against the point spread, but it seems not to account for it entirely.

A second possibility is attributional. Because predicting in line with one's intuition may "feel right," people may attribute successful intuitive (favorite) predictions to skill and unsuccessful favorite predictions to luck. Conversely, because predicting against one's intuitions may "feel wrong," people may attribute successful nonintuitive (underdog) predictions to luck and unsuccessful underdog predictions to skill. This attributional tendency could cause people to "learn" that predicting favorites is wiser than predicting underdogs, even if favorites lose more than underdogs against the spread.

At the end of the season, participants in the choice and warned choice conditions ($N = 75$) completed a survey designed to assess whether they attributed winning or losing predictions to luck or to skill. Approximately half of the participants were presented with a list of all of their losing predictions from weeks 14–16. For each prediction, they indicated whether they considered it a "Bad Decision" or whether they were incorrect because they were "Unlucky." The other half of the participants saw a list of their winning predictions from weeks 14–16. For each prediction, they indicated whether they considered it a "Good Decision" or whether they were correct because they were "Lucky."

The results of the choice and warned choice conditions were identical, and so we combined them for the analysis. A 2 (correct vs. incorrect prediction) × 2 (favorite vs. underdog prediction) ANOVA on the percentage of predictions participants attributed to luck yielded two major findings (see fig. 1). First, consistent with previous research (Gilovich 1983), participants were much more likely to attribute incorrect (vs. correct) predictions to luck ($F(1, 73) = 38.10$, $p < .001$). Second, and most important, there was a significant interaction ($F(1, 73) = 8.03$, $p < .006$), indicating that participants' attributions differed for favorite versus underdog predictions. Participants were significantly less likely to attribute correct favorite (vs. underdog) predictions to luck ($t(37) = 2.87$, $p = .007$), and they were somewhat more likely to attribute incorrect favorite (vs. underdog) predictions to luck ($t(36) = -1.21$, $p = .232$). These results are consistent with the attributional explanation for the increase in favorite predictions as the season progressed. Participants' attributional tendencies may have rendered them more likely to learn that a correct favorite (vs. underdog) prediction was good and that an incorrect underdog (vs. favorite) prediction was bad.

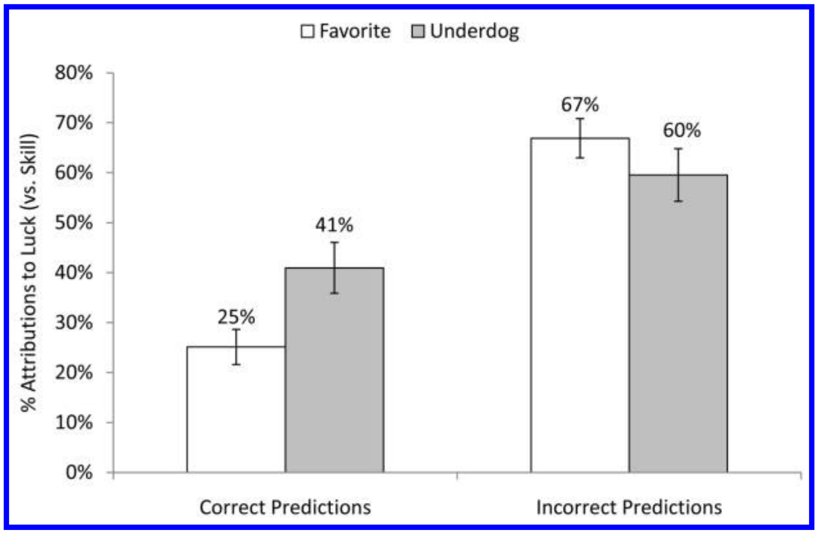
Can We Improve Crowd Wisdom by Altering the Way Predictions Are Elicited?

Although the crowd was systematically biased (and unwise) when *choosing* which team to bet on against point spreads, asking people to estimate the point differentials directly may cause them to focus on the very dimension (the point differential) that receives insufficient weight when making choices. Thus, this method of prediction elicitation may debias crowd ignorance, yielding less biased (and wiser) predictions. To test this hypothesis, we determined, for each game, the estimate condition's prediction against the spread. We did this by converting the mean (and median) point differential predictions into predictions against the point spread. Specifically, for each game we determined whether the mean (and median) predicted point differential was greater than, less than, or equal to the point spread presented in the choice conditions, indicating a prediction of "favorite," "underdog," or "no preference," respectively. (Note that using the median predicted point differential to determine the crowd's prediction is identical to first coding each prediction as one for or against the favorite, and then taking the majority's prediction as the crowd's prediction.)

Tables 2–4 display the results, and tables 2 and 3 also reveal the results of significance tests between conditions. The results of the estimate condition contrasted starkly with the results of the choice conditions. Using the mean predictions, the estimate condition crowd predicted the underdog in 82.7% of the games ($\chi^2(1, N = 226) = 96.92$, $p < .001$), correctly predicted 55.4% of the games against the spread ($\chi^2(1, N = 222) = 2.59$, $p = .107$), and outperformed 95.6% of its members. Using the median predictions, the estimate condition crowd predicted the underdog in the majority (70.4%) of the games in the sample ($\chi^2(1, N = 226) = 37.45$, $p < .001$), correctly predicted 50.9% of the games against the spread ($\chi^2(1, N = 214) = 0.07$, $p = .784$), and outperformed 57.8% of its individual members. (Neither prediction tendencies nor accuracy changed over time ($-.09 < r\text{'s} < .08$, $p\text{'s} > .180$).) Thus, different elicitation procedures yielded drastically different predictions ($\chi^2\text{s} > 177$, $p\text{'s} < .001$), and estimate condition predictions were wiser than choice condition predictions.

FIGURE 1

ATTRIBUTIONS OF PREDICTION SUCCESS TO LUCK VERSUS SKILL



The estimate condition crowd may have performed well either because estimating the point differential of the games yielded unbiased estimates or because estimating the point differential caused the crowd to exhibit a bias that is opposite the one exhibited by the choice condition crowds, namely, a bias toward the underdog. Only if the former is true should we conclude that estimating point differentials yields relatively wise predictions. To resolve this issue, we determined whether each prediction overestimated or underestimated the favorite's actual performance against the underdog. Using the mean prediction, the crowd overestimated the favorite's performance in 48.9% of the games. Using the median prediction, the crowd overestimated the favorite's performance in 52.4% of the games. Neither of these percentages differed significantly from 50%. Thus, the estimate condition's predictions were not biased toward underdogs (or favorites) but were equally likely to overestimate and underestimate the favorite's performance.

Having established that the estimate condition crowd predicted differently and more wisely than the choice condition crowd, we now attempt to explain this discrepancy. Analyzing the predictions of those in the choice/estimate condition—who predicted which team would win against the point spread before providing their point differential prediction—allows us to examine the merits of two alternative explanations. On the one hand, merely considering the favorite against a point spread may be enough to bias predictions toward favorites, perhaps because the point spread signals that the favorite is the better team. On the other hand, it may be the act of making point differential predictions that removes the bias toward favorites, perhaps because it encourages participants to consider the very dimension (the point differential) that they typically underweight when deciding which team to bet on. If the first explanation is true, then the choice/estimate condition's predictions should more closely resemble those of the choice conditions than those of the estimate condition. If the second explanation is true, then the choice/estimate condition's predictions should more closely resemble those of the estimate condition than those of the choice conditions.

The choice/estimate condition crowd was slightly more likely to predict favorites than was the estimate condition crowd, but it was dramatically less likely to predict favorites than the choice condition crowds (see table 2). Indeed, the choice/estimate condition crowd more closely resembled the estimate condition crowd than the choice condition crowds. Most notably, as was true of the estimate condition crowd, the choice/estimate condition crowd predicted underdogs more often than favorites (using the median predictions, χ^2 $(1, N = 218) = 3.60$, $p = .057$; using the mean predictions, χ^2 $(1, N = 222) = 61.39$, $p < .001$). All told, this suggests that, although considering the point spreads slightly increased the predictions of favorites, considering the point spreads did not induce an overall bias toward favorites, and thus that aspect of the method cannot fully explain the discrepancy between the choice and estimate conditions. Rather, it is the act of estimating the point differential (vs. choosing which team to bet on) that seems responsible for most of this discrepancy.

Why should estimating point differentials yield different predictions than choosing which team to bet on? We explore two possibilities, one contingent on a personal preference for winning teams (which we rule out) and one that emphasizes different psychological processes in choice and estimation (which we find support for).

First, we consider whether the discrepancy arises from the facts that (1) people bet on teams that they like more than teams that they dislike and (2) people tend to prefer good teams (favorites) to bad teams (underdogs). In order for choice condition participants to bet on a favored team that they like, they have to predict that the favorite will beat the spread. However, in order for estimate condition participants to bet on a favored team that they like, they have to predict only that the favorite will win the game—not that the favorite will beat the spread. Thus, when the favorite is preferred to the underdog, only estimate condition participants are able to bet on their preferred team without predicting that the team will beat the spread. Thus, predictions guided by preference may cause choice condition crowds to predict more favorites than estimate condition crowds.

To examine this possibility, we relied on data collected at the end-of-season survey, when we asked participants to rate their liking of each of the 32 NFL teams on a 7-point scale (-3 = strongly dislike; $+3$ = strongly like). For each prediction, we used the liking ratings to code whether the participant preferred the favorite, preferred the underdog, or had no preference between the two teams. Table 7 shows the percentage of favorites predicted by the crowd as a function of preference and condition. Consistent with the "liking" explanation, the crowd was more likely to predict favorites as their preference for the favorite increased. However, inconsistent with this explanation, the effect of liking on predictions was of equal size in the choice and estimate conditions. Indeed, table 7 shows that, no matter their preference, the choice condition crowds predicted more favorites than underdogs and the estimate condition crowd predicted more underdogs than favorites. Thus, although people do seem to bet on teams they like more than teams they dislike (Massey et al., forthcoming), this fact does not explain the discrepancy between the choice and estimate conditions.

Another possibility, suggested earlier, is that, although high intuitive confidence often causes bettors to underweight the point spread and to side with their intuitions (the favorite) when they are choosing which team to bet on (Simmons and Nelson 2006), asking bettors to predict the point differential of the game may attenuate the potentially biasing influence of intuitive confidence and cause them to weigh the point spread more heavily. To test this hypothesis, we examined condition differences in (1) the correlation between intuitive confidence—the percentage of people predicting the favorite to simply win the game in the Yahoo.com sample—and predictions and (2) the correlation between point spread magnitude and predictions. As shown in table 8, the correlations between intuitive confidence and the tendency to predict favorites were significantly higher in the two choice conditions (all r's > .38) than in the two estimate conditions (all r's < .19), with all z's > 2.20, p's < .029. Furthermore, the correlations between point spread magnitude and the tendency to predict favorites were significantly lower (i.e., more negative) in the two estimate conditions (all r's < $-.13$) than in the two choice conditions (all r's > .04), with all z's > 1.86, p's < .065 (15 out of the 16 differences in correlations were significant at p < .05). Thus, participants gave greater weight to intuitive confidence when they made choices and greater weight to the point spread's magnitude when they estimated point differentials directly.

TABLE 8

CORRELATIONS BETWEEN PREDICTING FAVORITES AGAINST THE SPREAD AND (1) INTUITIVE CONFIDENCE AND (2) POINT SPREAD MAGNITUDE

	Correlation with intuitive confidence (p-value)	Correlation with point spread magnitude (p-value)
Choice:		
Counting	.40	.05
	(< .001)	(.469)
Wager	.42	.08
	(< .001)	(.256)
Warned choice:		
Counting	.38	.05
	(< .001)	(.497)
Wager	.42	.05
	(< .001)	(.428)
Estimate:		
Median	.14	−.26
	(.036)	(< .001)
Mean	.19	−.13
	(.005)	(.048)
Choice/estimate:		
Median	.05	−.36
	(.503)	(< .001)
Mean	.15	−.19
	(.021)	(.005)

GENERAL DISCUSSION

Summary and Discussion

This research presented three major findings. First, when predicting against biased point spreads, crowds were systematically biased and ultimately unwise. This is a striking finding, especially because this investigation featured many elements that were amenable to the production of crowd wisdom. We investigated decision making in a domain that is widely believed to elicit crowd wisdom. We ensured that our sample included knowledgeable and enthusiastic football fans with diverse backgrounds and rooting interests. We gave all participants financial incentives to be accurate. We told a subset of the participants that the spreads were biased. And we conducted this study over the course of a 4-month-long season, allowing participants to learn over time. Despite these favorable elements, when predicting against point spreads, the crowd was systematically biased and consequently unwise.

This finding raises a number of questions. First, why are crowds systematically biased when predicting against point spreads? We believe that systematic biases arise because people are swayed by confidently held intuitions that favor the favorite and lead them to underweight point spreads (Simmons and Nelson 2006). Simmons and Nelson (2006) offered support for this theory in their investigation of predictions against point spreads, but the current investigation significantly extends this notion. First, we found that their theory applies not only to predictions against unbiased spreads but also to predictions against biased spreads. This is important because it means that the tendency to choose favorites over underdogs can be profitably exploited by increasing point spreads. Moreover, it suggests that their theory applies not only to situations that require participants to choose between equal alternatives but also to situations in which the nonintuitive choice is objectively and detectably superior.

Second, their theory also applies when people *know* that the spreads are biased. This is intriguing because it suggests that, in this context, the temptation to rely on one's intuitions is so strong as to lead people to rely on what they intuitively feel to be true (the favorite will prevail against the spread) rather than on what they generally know to be true (the favorite will usually lose against the spread). This finding is consistent with dual-process models of decision making, which emphasize how people often offer intuitive answers to questions even when they know, on a less emotional level, that those answers are inferior (Denes-Raj and Epstein 1994; Kahneman and Frederick 2002; Loewenstein et al. 2001; Shiv et al. 2005). It is also consistent with research showing that people often fail to apply general rules (e.g., base rates) to specific circumstances (Kahneman and Lovallo 1993; also see Dunning 2007). As John Steinbeck (1952, 448) wrote in *East of Eden*, "It is one of the triumphs of the human that he can know a thing and still not believe it."

Another question arising from this finding pertains to the operation of point spread betting markets. If gamblers are so heavily biased toward favorites, and if casinos know this (and they do; see Jeffries and Oliver 2000), then why do casinos not exploit this bias by increasing the spreads? One reason is that setting inaccurate spreads exposes the casino to risks that they would not face by setting accurate spreads. For example, although most people tend to bet on favorites, gamblers who rely on good mathematical models (rather than intuitive decision processes) to predict game outcomes could exploit inaccuracies in the spread. Indeed, there is anecdotal evidence that casinos are afraid of the "smart money," bettors who are ostensibly better than casinos at predicting a subset of game outcomes (Konik 2006). By aiming for accurate point spreads, casinos can ensure themselves of a long-run profit while dissuading the smart money from placing bets. A second reason has to do with competition. Especially in the age of the Internet, bettors often have many options when they are deciding which casino to use to place their bets, and bettors enticed to bet on favorites will prefer the casino offering the lowest spread. Thus, in order for a casino to attract most gamblers (i.e., those who tend to bet on favorites), the casino must offer point spreads that are no higher than the competition.

Our second major finding was that the choice condition crowds increasingly chose favorites over time even though choosing favorites produced worse outcomes. Thus, despite objective feedback to the contrary, these crowds actually seemed to learn that choosing favorites was wise. One possible and intriguing explanation for this result is attributional. Indeed, we discovered that people not only endorsed self-serving attributions, attributing winning predictions to skill and losing predictions to luck (Gilovich 1983), but they also endorsed *intuition-serving* attributions, as they were more likely to attribute intuitive (vs. nonintuitive) winning predictions to skill and intuitive (vs. nonintuitive) losing predictions to luck. Thus, in the long run, people are likely to reinforce themselves for intuitive predictions and punish themselves for nonintuitive predictions. This process may partially explain why strong intuitions may be resistant to change even in the face of objective feedback.

Finally, our third major finding is that, although crowd wisdom was absent from the (choice) conditions that predicted game outcomes against point spreads, predictions were drastically different—and wiser—among the (estimate) conditions that predicted the point differentials of the games. This difference may have emerged because estimating exact point differentials encourages people to give full weight to a dimension (the point spread) that is typically underweighted when they are *choosing* which team to bet on. This finding adds to a large literature showing that different methods of eliciting judgments induce different considerations and processes and hence often different judgments.

This finding suggests that, although crowds are unwise when the question is posed as it usually is in real-world betting contexts, a crowd randomly drawn from the same population may be wise when the question is posed differently. Thus, although this research suggests that wisdom-of-crowds proponents are wrong to assume that point spread betting markets offer evidence of crowd wisdom, crowd wisdom may emerge in (as of yet nonexistent) markets designed to elicit estimates of exact point differentials. This represents a generally important point about the elicitation of crowd wisdom. Although this research emphasizes that crowd wisdom may sometimes be elusive even under conditions of knowledge, motivation, independence, and diversity, it also emphasizes that crowd wisdom depends on the judgmental biases of the crowd members. Thus, predicting whether a crowd will be wise or unwise demands an understanding of the psychological processes induced by the judgment task. Although systematic biases may ruin the crowd's judgments when judgments are elicited in a manner that encourages intuitive responding, those biases may be absent from logically identical methods of eliciting the same information, and the crowd may emerge wiser.

Limitations and Future Directions

Any study demonstrating irrational behavior in the face of monetary incentives is left with an unanswered question, would the irrational behavior (betting on favorites) persist if the incentives were even larger or if participants faced the possibility of a monetary loss? This is a difficult question. On the one hand, it seems sensible to believe that increasing incentives would increase people's tendency to wisely choose underdogs, at least among those who know that favorites are disadvantaged. On the other hand, the evidence suggests that gamblers choose favorites not as a general strategy but because, for each individual game, their confidently held intuitions lead them to feel that the favorite is the correct choice. If gamblers truly feel, for each individual game, that the favorite is the correct choice, then it seems unlikely that increasing incentives would decrease their propensity to choose favorites, and it is even possible that increasing the stakes would increase reliance on intuition. This issue awaits empirical scrutiny.

Another unanswered question pertains to the effect of thought constraints (e.g., cognitive load, time pressure) on these predictions. According to our theory, such constraints should increase participants' tendency to choose in line with their intuitions, and we have some unpublished data supporting this. When asked to predict a week's worth of NFL games, time-constrained undergraduate participants chose favorites significantly more often (73.9%) than unconstrained participants (60.3%). (Please contact the first author for more details of this study.) The effect of thought constraints on estimating point differentials is less obvious. On the one hand, people may generate predicted point differentials by starting from a high anchor (the favorite will win by a lot) and adjusting downward, in which case time pressure would lead to insufficient adjustments that favor favorites more often. On the other hand, people may generate predicted point differentials by starting from a low anchor (e.g., one that starts at zero) and adjusting upward, in which case time pressure would lead to insufficient adjustments that favor favorites less often. Finally, it is possible that people generate these predictions not by an anchoring and adjustment process but by quickly retrieving and applying relevant knowledge from memory. In that case, we would expect time pressure to have no systematic effect on these judgments. We look forward to future research investigating this question.

Finally, it is interesting to consider the effect of expertise on these predictions. According to our theory, knowledgeable football fans will erroneously choose favorites against increased spreads because their intuitions so strongly signal that the favorite is the better team. If our study included people who knew nothing about NFL football, we would have expected them to lack strong intuitions about the favorite, to rely more on general rules about which team is the better bet, and to therefore choose more wisely against the spreads (especially if they were told that the spreads disadvantaged favorites). Thus, one counterintuitive prediction is that, because knowledge confers gamblers with misleading intuitions, those with greater (vs. less) knowledge about NFL football will be less wise when choosing against spreads. Interestingly, however, the effect of expertise would probably be quite different among those asked to estimate precise point differentials. In this case, intuitive biases that arise from increased knowledge seem not to interfere with the task, and so knowledge is likely to be beneficial—providing relevant information about the relative quality of the two teams—rather than harmful. We look forward to future research that tests this hypothesis.

REFERENCES

Alba, Joseph W. and J. Wesley Hutchinson (2000), "Knowledge Calibration: What Consumers Know and What They Think They Know," *Journal of Consumer Research*, 27 (September), 123–56.

American Gaming Association (2008), *State of the States: The AGA Survey of Casino Entertainment*, Washington, DC: American Gaming Association.

Avery, Christopher and Judith Chevalier (1999), "Identifying Investor Sentiment from Price Paths: The Case of Football Betting," *Journal of Business*, 72 (4), 493–521.

Bertrand, Marianne, Sendhil Mullainathan, and Eldar Shafir (2006), "Behavioral Economics and Marketing in Aid of Decision Making among the Poor," *Journal of Public Policy and Marketing*, 25 (1), 8–23.

Bettman, James R., Mary Frances Luce, and John W. Payne (1998), "Constructive Consumer Choice Processes," *Journal of Consumer Research*, 25 (2), 187–217.

Camerer, Colin F. (1998), "Can Asset Markets Be Manipulated? A Field Experiment with Racetrack Betting," *Journal of Political Economy*, 106 (3), 457–82.

Carmon, Ziv and Itamar Simonson (1998), "Price-Quality Trade-offs in Choice versus Matching: New Insights into the Prominence Effect," *Journal of Consumer Psychology*, 7 (4), 323–43.

Dana, James D., Jr. and Michael M. Knetter (1994), "Learning and Efficiency in a Gambling Market," *Management Science*, 40 (10), 1317–28.

Denes-Raj, Veronika and Seymour Epstein (1994), "Conflict between Intuitive and Rational Processing: When People Behave against Their Better Judgment," *Journal of Personality and Social Psychology*, 66 (5), 819–29.

Dunning, David (2007), "Prediction: The Inside View," in *Social Psychology: Handbook of Basic Principles*, 2nd ed., ed. Arie W. Kruglanski and E. Tory Higgins, New York: Guilford, 69–90.

Fischer, Gregory W., Ziv Carmon, Dan Ariely, and Gal Zauberman (1999), "Goal-Based Construction of Preferences: Task Goals and the Prominence Effect," *Management Science*, 45 (8), 1057–75.

Forsythe, Robert, Thomas A. Rietz, and Thomas W. Ross (1999), "Wishes, Expectations and Actions: A Survey on Price Formation in Election Stock Markets," *Journal of Economic Behavior and Organization*, 39 (1), 83–110.

Galton, Francis (1907), "Vox Populi," *Nature*, 75, 450–51.

Gandar, John, Richard Zuber, Thomas O'Brien, and Ben Russo (1988), "Testing Rationality in the Point Spread Betting Market," *Journal of Finance*, 43 (1), 995–1008.

Gilovich, Thomas (1983), "Biased Evaluation and Persistence in Gambling," *Journal of Personality and Social Psychology*, 44 (6), 1110–26.

Gilovich, Thomas, Dale Griffin, and Daniel Kahneman, eds. (2002), *Heuristics and Biases: The Psychology of Intuitive Judgment*, Cambridge: Cambridge University Press.

Gray, Philip K. and Stephen F. Gray (1997), "Testing Market Efficiency: Evidence from the NFL Sports Betting Market," *Journal of Finance*, 52 (4), 1725–37.

Grether, David M. and Charles R. Plott (1979), "Economic Theory of Choice and the Preference Reversal Phenomenon," *American Economic Review*, 69 (4), 623–38.

Hahn, Robert W. and Paul C. Tetlock, eds. (2006), *Information Markets: A New Way of Making Decisions*, Washington DC: AEI-Brookings Joint Center for Regulatory Studies.

Hastie, Reid and Tatsuya Kameda (2005), "The Robust Beauty of Majority Rules in Group Decisions," *Psychological Review*, 112 (2), 494–508.

Ho, Teck-Hua and Kay-Yut Chen (2007), "New Product Blockbusters: The Magic and Science of Prediction Markets," *California Management Review*, 50 (1), 144–58.

Hogarth, Robin M. (1978), "A Note on Aggregating Opinions," *Organizational Behavior and Human Performance*, 21 (1), 40–46.

Huber, Joel (1975), "Bootstrapping of Data and Decisions," *Journal of Consumer Research*, 2 (December), 229–34.

Jeffries, James and Charles Oliver (2000), *The Book on Bookies: An Inside Look at a Successful Sports Gambling Operation*, Boulder, CO: Paladin.

Kahneman, Daniel and Shane Frederick (2002), "Representativeness Revisited: Attribute Substitution in Intuitive Judgment," in *Heuristics and Biases: The Psychology of Intuitive Judgment*, ed. Thomas Gilovich, Dale Griffin, and Daniel Kahneman, Cambridge: Cambridge University Press, 49–81.

Kahneman, Daniel and Dan Lovallo (1993), "Timid Choices and Bold Forecasts: A Cognitive Perspective on Risk Taking," *Management Science*, 39 (1), 17–31.

Kahneman, Daniel and Amos Tversky, eds. (2000), *Choices, Values, and Frames*, New York: Russell Sage Foundation.

Knight, Hazel (1921), "Reliability of Judgments: A Comparison of Group and Individual," Master's thesis, Columbia University.

Konik, Michael (2006), *The Smart Money: How the World's Best Sports Bettors Beat the Bookies Out of Millions*, New York: Simon & Schuster.

Koriat, Asher (2008), "Subjective Confidence in One's Answers: The Consensuality Principle," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34 (4), 945–59.

Krizan, Zlatan and Paul D. Windschitl (2007), "The Influence of Outcome Desirability on Optimism," *Psychological Bulletin*, 133 (1), 95–121.

Kunda, Ziva (1990), "The Case for Motivated Reasoning," *Psychological Bulletin*, 108 (3), 480–98.

Larrick, Richard E. and Jack B. Soll (2006), "Intuitions about Combining Opinions: Misappreciation of the Averaging Principle," *Management Science*, 52 (1), 111–27.

Lee, Marcus and Gary Smith (2002), "Regression to the Mean and Football Wagers," *Journal of Behavioral Decision Making*, 15 (4), 329–42.

Levitt, Steven D. (2004), "Why Are Gambling Markets Organised So Differently from Financial Markets?" *Economic Journal*, 114 (April), 223–46.

List, John A. (2003), "Does Market Experience Eliminate Market Anomalies?" *Quarterly Journal of Economics*, 118 (1), 41–71.

Loewenstein, George F., Elke U. Weber, Christopher K. Hsee, and Ned Welch (2001), "Risk as Feelings," *Psychological Bulletin*, 127 (2), 267–86.

Lorge, Irving, David Fox, Joel Davitz, and Marlin Brenner (1958), "A Survey of Studies Contrasting the Quality of Group Performance and Individual Performance, 1920–1957," *Psychological Bulletin*, 55 (6), 337–72.

Makridakis, Spyros and Robert L. Winkler (1983), "Averages of Forecasts: Some Empirical Results," *Management Science*, 29 (9), 987–96.

Massey, Cade, Joseph P. Simmons, and David A. Armor (forthcoming), "Hope over Experience: Desirability and the Persistence of Optimism," *Psychological Science*.

Oliven, Kenneth and Thomas A. Rietz (2004), "Suckers Are Born but Markets Are Made: Individual Rationality, Arbitrage, and Market Efficiency on an Electronic Futures Market," *Management Science*, 50 (3), 336–51.

Oskarsson, An T., Leaf Van Boven, Gary H. McClelland, and Reid Hastie (2009), "What's Next? Judging Sequences of Binary Events," *Psychological Bulletin*, 135 (2), 262–85.

Radzevick, Joseph R. and Don A. Moore (2008), "Myopic Biases in Competitions," *Organization Behavior and Human Decision Processes*, 107 (2), 206–18.

Roxborough, Roxy and Mike Rhoden (1998), *Sports Book Management: A Guide for the Legal Bookmaker*, Las Vegas, NV: Las Vegas Sports Consultants.

Sauer, Raymond D. (1998), "The Economics of Wagering Markets," *Journal of Economic Literature*, 36 (December), 2021–64.

Shafir, Eldar and Robyn A. LeBoeuf (2002), "Rationality," *Annual Review of Psychology*, 53, 491–517.

Shiller, Robert J. (2005), *Irrational Exuberance*, 2nd ed., Princeton, NJ: Princeton University Press.

Shiv, Baba, George Loewenstein, Antoine Bechara, Hanna Damasio, and Antonio R. Damasio (2005), "Investment Behavior and the Negative Side of Emotion," *Psychological Science*, 16 (6), 435–39.

Simmons, Joseph P. and Leif D. Nelson (2006), "Intuitive Confidence: Choosing between Intuitive and Nonintuitive Alternatives," *Journal of Experimental Psychology: General*, 135 (3), 409–28.

Simonson, Itamar (1989), "Choice Based on Reasons: The Case of Attraction and Compromise Effects," *Journal of Consumer Research*, 16 (September), 158–74.

Slovic, Paul and Sarah Lichtenstein (1983), "Preference Reversals: A Broader Perspective," *American Economic Review*, 73 (4), 596–605.

Snowberg, Erik, Justin Wolfers, and Eric Zitzewitz (2005), "Information (In)Efficiency in Prediction Markets," in *Information Efficiency in Financial and Betting Markets*, ed. L. V. Williams, Cambridge: Cambridge University Press.

Soll, Jack B. and Richard P. Larrick (2009), "Strategies for Revising Judgment: How (and How Well) People Use Others' Opinions," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35 (3), 780–805.

Steinbeck, John (1952), *East of Eden*, New York: Viking.

Sunstein, Cass R. (2006), *Infotopia: How Many Minds Produce Knowledge*, New York: Oxford University Press.

Surowiecki, James (2004), *The Wisdom of Crowds*, New York: Doubleday.

Thaler, Richard H. (1985), "Mental Accounting and Consumer Choice," *Marketing Science*, 4 (3), 199–214.

Thaler, Richard H. and Cass R. Sunstein (2008), *Nudge: Improving Decisions about Health, Wealth, and Happiness*, New Haven, CT: Yale University Press.

Treynor, Jack L. (1987), "Market Efficiency and the Bean Jar Experiment," *Financial Analysts Journal*, 43 (May–June), 50–52.

Tversky, Amos, Shmuel Sattath, and Paul Slovic (1988), "Contingent Weighting in Judgment and Choice," *Psychological Review*, 95 (3), 371–84.

Tversky, Amos and Richard H. Thaler (1990), "Anomalies: Preference Reversals," *Journal of Economic Perspectives*, 4 (2), 201–11.

Yaniv, Ilan (2004), "The Benefits of Additional Opinions," *Current Directions in Psychological Science*, 13 (2), 1275–78.